Statistica Sinica **33** (2023), 1985-2016
doi:https://doi.org/10.5705/ss.202021.0087

PRINCIPAL COMPONENTS ANALYSIS FOR RIGHT CENSORED DATA

Benjamin W. Langworthy, Jianwen Cai, Robert W. Corty, Michael R. Kosorok and Jason P. Fine

University of North Carolina at Chapel Hill

Abstract: Principal components analysis (PCA) is a common dimension-reduction tool that transforms a set of variables into a linearly uncorrelated set of variables. Standard PCA estimators involve either the eigendecomposition of the estimated covariance matrix or a singular value decomposition of the centered data. However, for right-censored failure time data, estimating the principal components in this way is not straightforward because not all failure times are observed. Standard estimators for the covariance or correlation matrix should not be used in this case, because they require strong assumptions on the form of the joint distribution and on the marginal distributions beyond the final observation time. We present a novel, nonparametric estimator for the covariance of multivariate right-censored failure time data based on the counting processes and corresponding martingales defined by the failure times. We prove that these estimators are consistent and converge to a Gaussian process when properly standardized. We further show that these covariance estimates can be used to estimate a PCA for the martingales and counting processes for the different failure times. The corresponding estimates of the principal directions are consistent and asymptotically normal. We apply this method to data from a clinical trial of patients with pancreatic cancer, and recover a medically valid low-dimensional representation of adverse events.

Key words and phrases: Competing risks, multivariate survival analysis, principal components analysis.

1. Introduction

Principal components analysis (PCA), first introduced by Pearson (1901), transforms a set of potentially linearly correlated variables into a set of linearly uncorrelated variables, called the principal components. The transformations are all linear combinations of the data, and the first principal component is defined as the linear combination that explains the largest possible variance within the data. All subsequent directions are linear combinations of the data with maximal variance, subject to the constraint that they are orthogonal to all previous

Corresponding author: Benjamin W. Langworthy, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA. E-mail: langworthy.ben@gmail.com.

principal components. PCA is frequently used for dimension reduction and removing collinearity within a set of variables and, in general, is estimated using the eigendecomposition of the estimated covariance matrix or a singular value decomposition (SVD) of the mean centered data.

Although widely used, to the best of our knowledge, PCA has not been used for multivariate time-to-event data in the presence of right censoring. Such a setup can occur in longitudinal studies in which patients are followed for many different event types. We consider data from a clinical trial for patients with pancreatic cancer, but the techniques can be used for any longitudinal studies in which individuals are followed over time and data are collected on many different event types. We show that if there is right censoring for the different event types, either due to loss to follow up or an administrative end to study, standard PCA estimation techniques cannot be used for the time-to-event data. To show why this is, we consider a multivariate survival setting with p different event times. For each subject, we define $T^{(j)}$ as the failure time for the jth event type. If we assume that each subject also has an independent censoring time, C, the observed data for the jth event type consists of $Y^{(j)} = T^{(j)} \wedge C$, as well as an indicator for whether the observed time is a censoring time or a failure time. The full covariance matrix for $T = [T^{(1)}, \ldots, T^{(p)}]$ cannot be estimated without strong parametric assumptions. This means that a PCA cannot be estimated for T. In order to overcome this problem, we propose two versions of a PCA for multivariate survival data, and show how both can be estimated. One version uses the counting processes for each event type, defined as $N^{(j)}(t) = I(T^{(j)} \leq t)$. The second version uses the martingales based on the decomposition of these counting processes. Even though these counting processes are not observed for all time points if patients are censored, we show that the covariance between the counting processes and the martingales for two different event types can both be estimated nonparametrically. This allows us to consistently estimate the principal component directions for the counting processes and the martingales, even in the presence of independent censoring. The corresponding component scores can be estimated for each subject up until the time they are censored. In order to make these methods more flexible, we also allow for semi-competing risks, and show how to extend both survival PCA methods to this setting.

The estimation of the covariance between the counting processes and the martingales uses existing estimation methods for bivariate and univariate survival functions and for univariate hazard functions. Prentice and Cai (1992) show that at a fixed timepoint, the covariance for the counting process martingales for two different event types can be written as a function of bivariate and

univariate survival functions, and of univariate hazard functions. Similarly, the covariance between the counting processes for two different event types can be written as a function of the bivariate and univariate survival functions. We use the Kaplan–Meier and Nelson–Aalen estimators for the univariate survival and cumulative hazard functions. Estimators for the bivariate survival function include those of Dabrowska (1988), Prentice and Cai (1992), van der Laan (1993), and Lin and Ying (1993). If there are more than two event types, estimates of the full covariance and the full correlation matrices for the martingales or counting processes are found by estimating all the elements individually using bivariate methods, then we can estimate the principal component directions for the martingales and counting processes using the eigendecomposition of their estimated covariance or correlation matrices.

In the presence of competing risks, we use existing methods to estimate bivariate and univariate cause-specific hazard functions and cause-specific cumulative incidence functions (CIFs). Prentice et al. (1978) gives an overview of cause-specific hazard and incidence functions in the univariate setting, and Kalbfleisch and Prentice (2011) show how to estimate univariate cause-specific CIFs. Details on estimating bivariate cause-specific hazard functions and CIFs can be found in Cheng, Fine and Kosorok (2007). In order to extend our methods to the competing risk setting, we use the cause-specific counting processes and the martingales based on their decomposition. We show that the covariance between the cause-specific martingales or counting processes for two different events can be written as a function of the bivariate and univariate cause-specific CIFs and of the univariate cause-specific hazard functions. Using these results, we show that the full covariance matrix for cause-specific martingales and counting processes can be estimated in the presence of competing risks, which makes it possible to estimate the corresponding principal component directions using an eigendecomposition.

Being able to estimate the covariance and principal component directions for martingales and counting processes in the presence of competing risks is of particular interest in sick populations, where death acts as a competing risk for many adverse events. We present one such example using data from the metastatic pancreatic adenocarcinoma clinical trial (MPACT) study for patients with pancreatic cancer. In this study, patients are followed, and there are many types of adverse events, owing to both the cancer and the treatment. Using a PCA of the martingales, we define medically relevant groupings of the event types. We also show how the principal component scores can be estimated and used as predictors in a Cox proportional hazards (PH) model. This can be used to

remove the multicollinearity between predictors, and is analogous to a principal components linear regression.

The rest of this paper is structured as follows. Section 2 presents the estimation of covariance and correlation matrices for counting processes and martingales. Section 3 defines the estimators for a survival PCA, and shows that they are consistent and asymptotically normal. Section 4 provides the results of our simulation studies for survival PCA methods. Section 5 presents an analysis of adverse events among patients in the MPACT study for pancreatic cancer. Section 6 concludes the paper. Proofs for the theorems can be found in the appendix, and additional simulation results are provided in the Supplementary Material.

2. Covariance Estimation for Bivariate Counting Processes and Counting Process Martingales

2.1. Estimation of the covariance in the presence of right censoring

As in Section 1, we assume that for each subject in a population, there are p event types of interest, and let $T^{(j)}$ denote the failure time for the jth event type. We further assume that $T^{(j)}$ is a continuous random variable. The full vector of failure times for a subject can be written as $T = [T^{(1)}, \ldots, T^{(p)}]^T$. If $\mathbf{t} = [t_1, \ldots, t_p]^T$ is a vector of fixed timepoints, the joint distribution for T is defined as $F_T(t) = P(T^{(1)} \leq t_1, \ldots, T^{(p)} \leq t_p)$, and the univariate distribution functions are defined as $F^{(j)}(t_j) = P(T^{(j)} \leq t_j)$, for $j = 1, \ldots, p$. We also assume that there is an independent censoring time, C, with distribution $F_C(c) = P(C \leq c)$. This censoring time is the same for all event types. This assumption is reasonable when all event types are measured for the same subject and the end of the study or loss to follow up make it impossible to obtain any additional information from that subject for any of the event types. The observed data for the jth event type is the observed time, $Y^{(j)} = T^{(j)} \wedge C$, and the censoring indicator, $\eta^{(j)} = I(T^{(j)} \leq C)$. Define the counting process associated with the jth event type as $N^{(j)}(t) = I(T^{(j)} \leq t)$. Note that the value of $N^{(j)}(t)$ is not always observed, unlike $N^{(j1)}(t) = N^{(j)}(t)\eta^{(j)}$, which we use in later sections to derive the asymptotic properties of the estimates.

We define the cumulative hazard function for the jth event type at time t as $\Lambda^{(j)}(t) = \int_0^t \lambda^{(j)}(s)ds$, where $\lambda^{(j)}(t) = \lim_{\delta \to 0}(1/\delta)P(t \leq T^{(j)} < t + \delta | T^{(j)} \geq t)$. The martingale defined by the decomposition of $N^{(j)}(t)$ is $M^{(j)}(t) = N^{(j)}(t) - \Lambda^{(j)}(t \wedge T^{(j)})$.

Define the covariance between $N^{(j)}(t_j)$ and $N^{(j')}(t_{j'})$ as $CN^{(j,j')}(t_j, t_{j'})$ and, similarly, the covariance between $M^{(j)}(t_j)$ and $M^{(j')}(t_{j'})$ as $CM^{(j,j')}(t_j, t_{j'})$. Because $N^{(j)}(t)$ may not be observed for every subject in the presence of censoring, $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$ cannot be calculated using standard methods. Thus, we use the following equality to estimate $CN^{(j,j')}(t_j, t_{j'})$:

$$CN^{(j,j')}(t_j, t_{j'}) = S^{(j,j')}(t_j, t_{j'}) - S^{(j)}(t_j)S^{(j')}(t_{j'}), \tag{2.1}$$

where $S^{(j,j')}(t_j, t_{j'}) = P(T^{(j)} > t_j, T^{(j')} > t_{j'})$ is the bivariate survival function and $S^{(j)}(t_j) = P(T^{(j)} > t_j)$ is the univariate survival function. Similarly, the following equality from Prentice and Cai (1992) is obtained using Stieltjes integration and repeated integration by parts, and can be used to estimate $CM^{(j,j')}(t_j, t_{j'})$:

$$\begin{aligned} CM^{(j,j')}(t_j, t_{j'}) = S^{(j,j')}(t_j, t_{j'}) - 1 &+ \int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'})\lambda^{(j)}(s_j)ds \\ &+ \int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j'}^-)\lambda^{(j')}(s_{j'})ds_{j'} \\ &+ \int_0^{t_j}\int_0^{t_{j'}} S(s_j^-, s_{j'}^-)\lambda^{(j)}(s_j)\lambda^{(j')}(s_{j'})ds_j ds_{j'}. \end{aligned} \tag{2.2}$$

This allows for consistent estimations of $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$ by plugging in consistent estimates of all of the quantities on the right-hand side of Equations (2.1) and (2.2), respectively. The variances of $N^{(j)}(t)$ and $M^{(j)}(t)$ can also be written as functions of the univariate survival functions, $CN^{(j,j)}(t_j) = S^{(j)}(t_j)\{1 - S^{(j)}(t_j)\}$, and $CM^{(j,j)}(t_j) = 1 - S^{(j)}(t_j)$, respectively. We use the Kaplan–Meier estimator for the univariate survival functions, and the Nelson–Aalen estimator for the univariate cumulative hazard functions. Potential estimators for the bivariate survival function are discussed in Section 1. We use the estimator from Dabrowska (1988), because it is shown to converge weakly to a Gaussian process (Gill, van der Laan and Wellner (1995)) and performs well in simulations (Cheng, Fine and Kosorok (2007)).

In addition to the covariance, the correlation can be a useful measure for describing the relationship between the martingales or counting processes of two event types. The correlation is standardized by the product of the standard deviations of the martingales or counting processes for the two event types. This gives a useful way of comparing the strength of the linear association of two martingales or counting processes across time in a way that is not influenced by the changing variance of the martingales and counting processes over time. Define $RN^{(j,j')}(t_j, t_{j'}) = Cor\{N^{(j)}(t_j), N^{(j')}(t_{j'})\}$ and $RM^{(j,j')}(t_j, t_{j'}) =$

$Cor\{M^{(j)}(t_j), M^{(j')}(t_{j'})\}$. The following equality can be used to estimate $RN^{(j,j')}(t_j, t_{j'})$:

$$RN^{(j,j')}(t_j, t_{j'}) = \frac{CN^{(j,j')}(t_j, t_{j'})}{\sqrt{S^{(j)}(t_j)\{1 - S^{(j)}(t_j)\}}\sqrt{S^{(j')}(t_{j'})\{1 - S^{(j')}(t_{j'})\}}}. \quad (2.3)$$

Similarly, the following equality can be used to estimate $RM^{(j,j')}(t_j, t_{j'})$:

$$RM^{(j,j')}(t_j, t_{j'}) = \frac{CM^{(j,j')}(t_j, t_{j'})}{\sqrt{1 - S^{(j)}(t_j)}\sqrt{1 - S^{(j')}(t_{j'})}}. \quad (2.4)$$

Note that Equation (2.3) requires that $0 < S^{(j)}(t_j) < 1$, for all j, and Equation (2.4) requires that $S^{(j)}(t_j) < 1$, for all j, in order to be well defined. The right-hand sides of Equations (2.3) and (2.4) can be estimated consistently using the Kaplan–Meier estimator and estimates for $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$, respectively.

In addition, note that $N^{(j)}(t)$ contains information only on whether an event has happened before time t, and no further information on when it happened. This means $CN^{(j,j')}(t_j, t_{j'})$ and $RN^{(j,j')}(t_j, t_{j'})$ are only useful when $0 < S^{(j)}(t_j), S^{(j')}(t_j) < 1$. For example, when t_j or $t_{j'} = \infty$, $CN^{(j,j')}(t_j, t_{j'}) = 0$ and $RN^{(j,j')}(t_j, t_{j'})$ is undefined, because either $N^{(j)}(t_j)$ or $N^{(j')}(t_{j'})$ will be a degenerate random variable. As an alternative, $M^{(j)}(t)$ contains information both on whether an event has happened by time t and when it happened, if it happened before time t. If we evaluate $CM^{(j,j')}(t_j, t_{j'})$ or $RM^{(j,j')}(t_j, t_{j'})$ at $t_j = t_{j'} = \infty$, they are equivalent to the covariance and correlation, respectively, between $T^{(j)}$ and $T^{(j')}$ after transforming both variables to have an Exponential(1) distribution using a cumulative hazard transformation. That is, $CM^{(j,j')}(\infty, \infty) = \text{Cov}\{\Lambda^{(j)}(T^{(j)}), \Lambda^{(j')}(T^{(j)})\}$ and $RM^{(j,j')}(\infty, \infty) = \text{Cor}\{\Lambda^{(j)}(T^{(j)}), \Lambda^{(j')}(T^{(j)})\}$. However, this cumulative hazard transformation, and other similar transformations, such as the CDF transformation, are not possible for all observations in the presence of right censoring. This is why we focus on martingale and counting processes, which have correlation and covariance functions indexed by time and can be estimated in the presence of right censoring.

For a fixed set of timepoints, $\mathbf{t} = [t_1, \ldots, t_p]^T$, we define the full covariance matrix for all p counting processes as

$$CN(\mathbf{t}) = \begin{bmatrix} CN^{(1,1)}(t_1,t_1) & \cdots & CN^{(1,p)}(t_1,t_p) \\ \vdots & \ddots & \vdots \\ CN^{(p,1)}(t_p,t_1) & \cdots & CN^{(p,p)}(t_p,t_p) \end{bmatrix}. \tag{2.5}$$

Here, $RN(\mathbf{t})$, $CM(\mathbf{t})$, and $RM(\mathbf{t})$ can be defined analogously. We denote the estimators of each of these matrices created using the estimator for each element of the matrix as $\widehat{CN}(\mathbf{t})$, $\widehat{RN}(\mathbf{t})$, $\widehat{CM}(\mathbf{t})$, and $\widehat{RM}(\mathbf{t})$ respectively. Note that these estimates may not be positive semidefinite. In this case several methods can be used to transform these matrices to be positive semidefinite or positive definite (Rousseeuw and Molenberghs (1993)). We define $\widetilde{CN}(\mathbf{t})$ as the matrix with the same eigenvectors as $\widehat{CN}(\mathbf{t})$, and with all negative eigenvalues set to some small nonnegative constant. Then, $\widetilde{RN}(\mathbf{t})$, $\widetilde{CM}(\mathbf{t})$, and $\widetilde{RM}(\mathbf{t})$ are defined analogously. If we assume that the true covariance or correlation matrices are positive definite, then it can be shown that transforming these matrices to be positive definite does not change the limiting behavior at a fixed timepoint, using results from Section 2.3 and Weyl's inequality.

In the presence of censoring, the values of $\mathbf{t}$ for which the martingale or counting process covariance and correlation matrices should be estimated depends on the largest observed censoring or failure time for each event type. For a given data set, define t_j^* as the largest observed failure or censoring time for the jth event type. The Kaplan–Meier estimator for the survival function for event type j is typically not estimated beyond timepoint t_j^*, because doing so can introduce bias (Gillespie, Gillespie and Iglewicz (1992)). The Nelson–Aalen estimator and nonparametric bivariate survival function estimators also have this problem. For this reason, we recommend only the estimations of $CM^{(j,j')}(t_j,t_{j'})$, $RM^{(j,j')}(t_j,t_{j'})$, $CN^{(j,j')}(t_j,t_{j'})$, and $RN^{(j,j')}(t_j,t_{j'})$, for $t_j \le t_j^*$ and $t_{j'} \le t_{j'}^*$. Furthermore, when the final observed time for the jth event type is a failure time, then $\hat{S}^{(j)}(t_j) = \hat{S}^{(j,j')}(t_j,t_{j'}) = 0$, for $t_j \ge t_j^*$. This means that $\widehat{CN}^{(j,j')}(t_j,t_{j'}) = 0$ and $\widehat{RN}^{(j,j')}(t_j,t_{j'})$ is not well defined for $t_j \ge t_{j'}$. A similar problem happens before the first observed failure for event type j when $\hat{S}^{(j)}(t_j) = 1$. In this case, neither $\widehat{RN}^{(j,j')}(t_j,t_{j'})$ nor $\widehat{RM}^{(j,j')}(t_j,t_{j'})$ is well defined. In general, the martingale and counting process covariances and correlations should only be estimated for timepoints between the first observed failure time and the final observed failure or censoring time for each event type.

2.2. Estimation of the covariance in the presence of right censoring and competing risks

In this section, we allow for the introduction of a single competing risk for all of the noncompeting event types of interest. As a motivating example, consider a study in which subjects are followed over time for a number of noncompeting adverse events, and death acts as a competing risk for each of the adverse events. As before, we assume that each subject has p noncompeting failure times, $T = [T^{(1)}, \ldots, T^{(p)}]^T$. Define $\ddot{T}$ as the failure time for the competing risk. If there is more than one competing risk, these can be combined into a single competing event time (Cheng, Fine and Kosorok (2007)). In the absence of censoring for the jth event type, we observe $\ddot{T}^{(j)} = T^{(j)} \wedge \ddot{T}$ and $\ddot{\gamma}^{(j)} = 2 - I(\ddot{T} > T^{(j)})$. If we again assume an independent censoring time, C, we observe $\ddot{Y}^{(j)} = \ddot{T}^{(j)} \wedge C$, the failure indicator $\ddot{\eta}^{(j)} = I(\ddot{T}^{(j)} \leq C)$, and cause and censoring-type indicator, $\ddot{\epsilon}^{(j)} = \ddot{\eta}^{(j)}\ddot{\gamma}^{(j)}$. This setup can be thought of as a semi-competing risk setting, where $T^{(j)}$ is not observed if $\ddot{T}$ is observed first, but $\ddot{T}$ may still be observed after $T^{(j)}$ is observed. Previous studies on semi-competing risks include those of Fine, Jiang and Chappell (2001), Cheng, Fine and Kosorok (2007), and Jazić et al. (2016). Importantly, as with Cheng, Fine and Kosorok (2007), the event types in T do not compete with each other, and it is the association between these noncompeting event types that we focus on in this section.

The cause-specific counting process for the jth event type and lth cause is defined as $\ddot{N}_l^{(j)}(t) = I(\ddot{T}^{(j)} \leq t, \ddot{\gamma}^{(j)} = l)$, for $j = 1, \ldots, p$ and $l = 1, 2$. Similarly to the previous section, $\ddot{N}_l^{(j)}$ is not always observed, and so we also define $\ddot{N}_l^{(j1)}(t) = \ddot{N}_l^{(j)}(t)\ddot{\eta}^{(j)}$ for estimation purposes. The cumulative cause-specific hazard for the jth event type and lth cause evaluated at time t is $\ddot{\Lambda}_l^{(j)}(t) = \int_0^t \ddot{\lambda}_l^{(j)}(s)ds$, where $\ddot{\lambda}_l^{(j)}(t) = \lim_{\delta \to 0}(1/\delta)P(t \leq \ddot{T}^{(j)} < t + \delta, \ddot{\gamma}^{(j)} = l | \ddot{T}^{(j)} \geq t)$. The martingale based on the decomposition of $\ddot{N}_l^{(j)}(t)$ is $\ddot{M}_l^{(j)}(t) = \ddot{N}_l^{(j)}(t) - \ddot{\Lambda}_l^{(j)}(t \wedge \ddot{T}^{(j)})$ (Lin (1997)).

We focus on the covariance between the counting processes and the martingales for the noncompeting adverse events. That is, $\ddot{CN}^{(j,j')}(t_j, t_{j'}) = Cov\{\ddot{N}_1^{(j)}(t_j), \ddot{N}_1^{(j')}(t_{j'})\}$ and $\ddot{CM}^{(j,j')}(t_j, t_{j'}) = Cov\{\ddot{M}_1^{(j)}(t_j), \ddot{M}_1^{(j')}(t_{j'})\}$. These quantities cannot be estimated using standard methods in the presence of censoring. The following equality can be used to estimate $\ddot{CN}^{(j,j')}(t_j, t_{j'})$:

$$\ddot{CN}^{(j,j')}(t_j, t_{j'}) = F_{11}^{(j,j')}(t_j, t_{j'}) - F_1^{(j)}(t_j)F_1^{(j')}(t_{j'}), \tag{2.6}$$

where $F_{kl}^{(j,j')}(t_j, t_{j'}) = P(\ddot{T}^{(j)} \leq t_j, \ddot{\gamma}^{(j)} = k, \ddot{T}^{(j')} \leq t_{j'}, \ddot{\gamma}^{(j')} = l)$ is the bivariate cause-specific CIF and $F_k^{(j)}(t_j) = P(\ddot{T}^{(j)} \leq t_j, \ddot{\gamma}^{(j)} = k)$ is the univariate

cause-specific CIF. Note that in our setup $F_{22}^{(j,j')}(t,t)$ simplifies to a univariate function, because the competing risk time is the same for event types j and j'. For notational purposes, for a bivariate function, $G(t_1,t_2)$, define $G(dt_1,dt_2) = G^{(1,1)}(t_1,t_2)dt_1dt_2$, $G(dt_1,t_2) = G^{(1,0)}(t_1,t_2)dt_1$, and $G(t_1,dt_2) = G^{(0,1)}(t_1,t_2)dt_2$, where $G^{(1,1)}(t_1,t_2)$ is the second partial derivative of $G(t_1,t_2)$ with respect to t_1 and t_2, $G^{(1,0)}(t_1,t_2)$ is the first partial derivative of $G(t_1,t_2)$ with respect to t_1, and $G^{(0,1)}(t_1,t_2)$ is the first partial derivative of $G(t_1,t_2)$ with respect to t_2. Using the fact that $P(\ddot{T}_j \le s, \ddot{\gamma}_j = k, \ddot{T}_{j'} > t) = F_k^{(j)}(s) - F_{k1}^{(j,j')}(s,t) - F_{k2}^{(j,j')}(s,t)$, the following result is obtained and can be used to estimate $C\ddot{M}^{(j,j')}(t,t)$:

$$
\begin{aligned}
&C\ddot{M}^{(j,j')}(t,t) = \\
&\ddot{\Lambda}_1^{(j)}(t)\ddot{\Lambda}_1^{(j')}(t)\ddot{S}^{(j,j')}(t,t) + \int_0^t \ddot{\Lambda}_1^{(j)}(s)\ddot{\Lambda}_1^{(j')}(s)F_{22}^{(j,j')}(ds) \\
&+ \int_0^t \int_0^t \{1-\ddot{\Lambda}_1^{(j)}(s_1)\}\{1-\ddot{\Lambda}_1^{(j')}(s_2)\}F_{11}^{(j,j')}(ds_1,ds_2) \\
&+ \int_0^t \{1-\ddot{\Lambda}_1^{(j)}(s)\}\{-\ddot{\Lambda}_1^{(j')}(t)\}\{F_1^{(j)}(ds) - F_{11}^{(j,j')}(ds,t) - F_{12}^{(j,j')}(ds,t)\} \\
&+ \int_0^t \{1-\ddot{\Lambda}_1^{(j')}(s)\}\{-\ddot{\Lambda}_1^{(j)}(t)\}\{F_1^{(j')}(ds) - F_{11}^{(j,j')}(t,ds) - F_{21}^{(j,j')}(t,ds)\} \\
&+ \int_0^t \int_{s_1}^t \{1-\ddot{\Lambda}_1^{(j)}(s_1)\}\{-\ddot{\Lambda}_1^{(j')}(s_2)\}F_{12}^{(j,j')}(ds_1,ds_2) \\
&+ \int_0^t \int_{s_1}^t \{1-\ddot{\Lambda}_1^{(j')}(s_1)\}\{-\ddot{\Lambda}_1^{(j)}(s_2)\}F_{12}^{(j',j)}(ds_1,ds_2), \qquad (2.7)
\end{aligned}
$$

where $\ddot{S}^{(j,j')}(t_j,t_{j'}) = P(\ddot{T}^{(j)} > t_j, \ddot{T}^{(j')} > t_{j'})$. Equation (2.7) is obtained using Stieltjes integration, similarly to how Equation (2.2) is obtained in Prentice and Cai (1992). The bivariate cause-specific CIFs can be estimated using methods from Cheng, Fine and Kosorok (2007), and the univariate cause-specific CIFs and cumulative hazards can be estimated using methods from Kalbfleisch and Prentice (2011). Both the bivariate CIF and $\ddot{S}^{(j,j')}(t_j,t_{j'})$ require estimating a standard noncompeting risk bivariate survival function. Here, we use the Dabrowska estimator, as in Section 2.1. Therefore, the estimators for $C\ddot{N}^{(j,j')}(t_j,t_{j'})$ and $C\ddot{M}^{(j,j')}(t_j,t_{j'})$ are obtained by consistently estimating all the elements on the right-hand side of Equations (2.6) and (2.7), respectively. The variances of $\ddot{N}^{(j)}(t_j)$ and $\ddot{M}^{(j)}(t_j)$ can also be written as functions of the bivariate and univariate cause-specific CIFs, $C\ddot{N}^{(j,j)}(t_j) = F_1^{(j)}(t_j)\{1-F_1^{(j)}(t_j)\}$ and $C\ddot{M}^{(j,j)}(t_j) = F_1^{(j)}(t_j)$, respectively.

As in the previous section, we can extend these methods to estimate the correlation between the counting processes or martingales. Define $\ddot{RN}^{(j,j)}(t_j, t_{j'}) = Cor\{\ddot{N}_1^{(j)}(t_j), \ddot{N}_1^{(j')}(t_{j'})\}$ and $\ddot{RM}^{(j,j)}(t_j, t_{j'}) = Cor\{\ddot{M}_1^{(j)}(t_j), \ddot{M}_1^{(j')}(t_{j'})\}$. The following equality is used to estimate $\ddot{RN}^{(j,j)}(t_j, t_{j'})$:

$$\ddot{RN}^{(j,j)}(t_j, t_{j'}) = \frac{\ddot{CN}^{(j,j')}(t_j, t_{j'})}{\sqrt{F_1^{(j)}(t_j)\{1 - F_1^{(j)}(t_j)\}}\sqrt{F_1^{(j')}(t_{j'})\{1 - F_1^{(j')}(t_{j'})\}}}, \tag{2.8}$$

and

$$\ddot{RM}^{(j,j)}(t) = \frac{\ddot{CM}^{(j,j')}(t)}{\sqrt{F_1^{(j)}(t)}\sqrt{F_1^{(j')}(t)}} \tag{2.9}$$

can be used to estimate $\ddot{RM}^{(j,j)}(t_j, t_{j'})$. Equation (2.8) requires $0 < F_1^{(j)}(t_j) < 1$, for all j, and Equation (2.9) requires $0 < F_1^{(j)}(t)$, for all j, in order to be well defined. The full covariance and correlation matrices at the vector of timepoints $\mathbf{t}$, $\ddot{CN}(\mathbf{t}), \ddot{RN}(\mathbf{t}), \ddot{CM}(\mathbf{t})$, and $\ddot{RM}(\mathbf{t})$, can be defined analogously to $CN(\mathbf{t})$, $RN(\mathbf{t})$, $CM(\mathbf{t})$, and $RM(\mathbf{t})$, respectively, from the previous section. The estimators of these matrices created by estimating each element of the matrix are defined as $\widehat{\ddot{CN}}(\mathbf{t}), \widehat{\ddot{RN}}(\mathbf{t}), \widehat{\ddot{CM}}(\mathbf{t})$, and $\widehat{\ddot{RM}}(\mathbf{t})$, respectively. These matrix estimates may not be positive definite. In this case, we can define $\widetilde{\ddot{CN}}(\mathbf{t}), \widetilde{\ddot{RN}}(\mathbf{t}), \widetilde{\ddot{CM}}(\mathbf{t})$, and $\widetilde{\ddot{RM}}(\mathbf{t})$ using the same techniques as in Section 2.1. As in Section 2.1, the values of $\mathbf{t}$ for which it is useful to estimate $\ddot{CN}(\mathbf{t}), \ddot{RN}(\mathbf{t}), \ddot{CM}(\mathbf{t})$, and $\ddot{RM}(\mathbf{t})$ are meaningful depend on the specific data set. It is not useful to consider values of t_j larger than the largest observed failure, censoring, or competing risk time for event type j, because this will introduce bias. For values of t_j smaller than the first observed event time for event type j, the estimates of the competing risk martingale and counting process correlation matrices are not well defined, because the estimate of $F_1^{(j)}(t_j)$ is zero.

2.3. Weak convergence of the covariance and correlation estimates

In this section, we show the weak convergence properties for the estimates of the covariance and correlation matrices for the martingales and the counting processes. Define PF as the expectation of a random function, F, and $H^{(j)}(t) = I(Y^{(j)} \geq t)$. Furthermore, define $\mathbb{P}_n N^{(j1)}(t) = (1/n)\sum_{i=1}^n I(Y_i^{(j)} \leq t, \eta_i^{(j)} = 1)$ and $\mathbb{P}_n H^{(j)}(t) = (1/n)\sum_{i=1}^n I(Y_i^{(j)} \geq t)$, and for an arbitrary $q \times r$ matrix M, define the function $\mathrm{Vec}(M)$ as the column vector created by stacking the columns of M. In general, we assume that $\Lambda^{(j)}$ is estimated using the Nelson–Aalen

estimator, and $S^{(j)}$ is estimated using the Kaplan–Meier estimator. Theorem 1 shows the weak convergence of the estimates of the elements of CM, CN, RM, and RN.

Theorem 1. *Assume that the estimator of $S^{(j,j')}$ converges weakly, such that $\sqrt{n}[\mathbb{P}_n N^{(11)} - PN^{(11)}, \ldots, \mathbb{P}_n N^{(p1)} - PN^{(p1)}, \mathbb{P}_n H^{(1)} - PH^{(1)}, \ldots, \mathbb{P}_n H^{(p)} - PH^{(p)}, \hat{S}^{(1,2)} - S^{(1,2)}, \ldots, \hat{S}^{(p-1,p)} - S^{(p-1,p)}]^T \rightsquigarrow [Z_{N_1}, \ldots, Z_{N_p}, Z_{H_1}, \ldots, Z_{H_p}, Z_{S_{12}}, \ldots, Z_{S_{p-1,p}}]^T$ in $D[0,\tau_1]^2 \times \cdots \times D[0,\tau_p]^2 \times D[0,\tau_{12}] \times \cdots \times D[0,\tau_{p-1,p}]$, where $\tau_{jj'} = (\tau_j, \tau_{j'})$, $(D[0,\tau_j], ||\cdot||_\infty)$ is the space of univariate cadlag functions of bounded variation in $[0,\tau_j]$ equipped with a uniform norm, $(D[0,\tau_{jj'}], ||\cdot||_\infty)$ is the space of bivariate cadlag functions of bounded variation in $[0,\tau_{jj'}]$ equipped with a uniform norm, and $[Z_{N_1}, \ldots, Z_{N_p}, Z_{H_1}, \ldots, Z_{H_p}, Z_{S_{12}}, \ldots, Z_{S_{p-1,p}}]^T$ is a mean-zero tight Gaussian process. Assume that $P(Y^{(j)} > \tau_j, Y^{(j')} > \tau_{j'}) > 0$ for all $1 \le j, j' \le p$, and $\Lambda^{(j)} < \infty$ for $j = 1, \ldots, p$. Then for any $[0,\mathbf{t}] \subset [0,\tau]$, where $\tau = [\tau_1, \ldots, \tau_p]$,*

$$\sqrt{n}[\mathit{Vec}(\widehat{CM}) - \mathit{Vec}(CM)](\mathbf{t}) \rightsquigarrow Z_{CM} \tag{2.10}$$

$$\sqrt{n}[\mathit{Vec}(\widehat{CN}) - \mathit{Vec}(CN)](\mathbf{t}) \rightsquigarrow Z_{CN}, \tag{2.11}$$

where Z_{CM} and Z_{CN} are p^2-dimensional mean-zero Gaussian processes. Furthermore, if $\omega = [\omega_1, \ldots, \omega_p]$, where $\omega_j < \tau_j$ for $j = 1, \ldots, p$ and $P(Y^{(j)} \le \omega_j, Y^{(j')} \le \omega_{j'}) > 0$ for all $1 \le j, j' \le p$, then for any $[\omega,\mathbf{t}] \subset [\omega,\tau]$,

$$\sqrt{n}[\mathit{Vec}(\widehat{RM}) - \mathit{Vec}(RM)](\mathbf{t}) \rightsquigarrow Z_{RM} \tag{2.12}$$

$$\sqrt{n}[\mathit{Vec}(\widehat{RN}) - \mathit{Vec}(RN)](\mathbf{t}) \rightsquigarrow Z_{RN}, \tag{2.13}$$

where Z_{RM} and Z_{RN} are p^2-dimensional mean-zero Gaussian processes.

The proof for this theorem is presented in the Appendix, and is done by showing that the right-hand sides of Equations (2.11) and (2.10) are Hadamard differentiable mappings, and then applying the functional delta method (Theorem 2.8 in Kosorok (2008)). The additional assumption required for Equations (2.12) and (2.13) is needed to ensure that $RM^{(j,j)}$ and $RN^{(j,j')}$ are well defined. In addition, the assumption of joint weak convergence can be shown for the Nelson–Aalen, Kaplan–Meier and Dabrowska estimators using methods similar to those for Lemma A.6 in the appendix of Cheng, Fine and Kosorok (2007). Next, we show the weak convergence properties for the martingale and counting process covariance and correlation matrices in the presence of competing risks. Define $\ddot{H}^{(j)}(t) = I(\ddot{Y}^{(j)} \ge t)$. Furthermore, let $\mathbb{P}_n \ddot{N}_l^{(j1)}$ and $\mathbb{P}_n \ddot{H}^{(j)}$ be defined analogously to $\mathbb{P}_n N^{(j1)}$ and $\mathbb{P}_n H^{(j)}$, respectively. We assume that $\ddot{\Lambda}_1^{(j)}$ is estimated using a Nelson–Aalen-style estimator, $F_1^{(j)}$ is estimated using methods

from Kalbfleisch and Prentice (2011), and $F_{kl}^{(j,j')}$ is estimated using methods from Cheng, Fine and Kosorok (2007). This leads to Theorem 2, which shows the weak convergence for the estimates of the elements of $\ddot{CM}$, $\ddot{CN}$, $\ddot{RM}$, and $\ddot{RN}$.

Theorem 2. *Assume that the estimator of* $\ddot{S}^{(j,j')}$ *converges weakly, such that* $\sqrt{n}[\mathbb{P}_n\ddot{N}_1^{(11)} - P\ddot{N}_1^{(11)}, \ldots, \mathbb{P}_n\ddot{N}_1^{(p1)} - P\ddot{N}_1^{(p1)}, \mathbb{P}_n\ddot{H}^{(1)} - P\ddot{H}^{(1)}, \ldots, \mathbb{P}_n\ddot{H}^{(p)} - P\ddot{H}^{(p)}, \hat{\ddot{S}}^{(1,2)} - \ddot{S}^{(1,2)}, \ldots, \hat{\ddot{S}}^{(p-1,p)} - \ddot{S}^{(p-1,p)}]^T \rightsquigarrow [Z_{\ddot{N}_1}, \ldots, Z_{\ddot{N}_p}, Z_{\ddot{H}_1}, \ldots, Z_{\ddot{H}_p}, Z_{\ddot{S}_{12}}, \ldots, Z_{\ddot{S}_{p-1,p}}]^T$ *in* $D[0,\ddot{\tau}_1]^2 \times \cdots \times D[0,\ddot{\tau}_p]^2 \times D[0,\ddot{\tau}_{12}] \times \cdots \times D[0,\ddot{\tau}_{p-1,p}]$, *where* $\ddot{\tau}_{jj'} = (\ddot{\tau}_j, \ddot{\tau}_{j'})$, $(D[0,\ddot{\tau}_j], ||\cdot||_\infty)$ *is the space of univariate cadlag functions of bounded variation in* $[0,\ddot{\tau}_j]$ *equipped with a uniform norm,* $(D[0,\ddot{\tau}_{jj'}], ||\cdot||_\infty)$ *is the space of bivariate cadlag functions of bounded variation in* $[0,\ddot{\tau}_{jj'}]$ *equipped with a uniform norm, and* $[Z_{\ddot{N}_1}, \ldots, Z_{\ddot{N}_p}, Z_{\ddot{H}_1}, \ldots, Z_{\ddot{H}_p}, Z_{\ddot{S}_{12}}, \ldots, Z_{\ddot{S}_{p-1,p}}]^T$ *is a mean-zero tight Gaussian process. Assume that* $P(\ddot{Y}^{(j)} > \tau_j, \ddot{Y}^{(j')} > \tau_{j'}) > 0$ *for all* $1 \le j, j' \le p$, *and* $\ddot{\Lambda}_k^{(j)} < \infty$ *for* $j = 1, \ldots, p$ *and* $k = 1, 2$. *Then, for any* $[0,\ddot{t}] \subset [0,\ddot{\tau}]$, *where* $\ddot{\tau} = [\ddot{\tau}_1, \ldots, \ddot{\tau}_p]$,

$$\sqrt{n}[\,Vec(\widehat{\ddot{CM}}) - Vec(\ddot{CM})](t) \rightsquigarrow Z_{\ddot{CM}} \tag{2.14}$$

$$\sqrt{n}[\,Vec(\widehat{\ddot{CN}}) - Vec(\ddot{CN})](t) \rightsquigarrow Z_{\ddot{CN}}, \tag{2.15}$$

where $Z_{\ddot{CM}}$ *and* $Z_{\ddot{CN}}$ *are* p^2*-dimensional mean-zero Gaussian processes. In addition, if* $\ddot{\omega} = [\ddot{\omega}_1, \ldots, \ddot{\omega}_p]$, *where* $\ddot{\omega}_j < \ddot{\tau}_j$ *for* $j = 1, \ldots, p$ *and* $P(\ddot{Y}^{(j)} \le \ddot{\omega}_j, \ddot{Y}^{(j')} \le \ddot{\omega}_{j'}) > 0$ *for all* $1 \le j, j' \le p$, *then for any* $[\ddot{\omega},\ddot{t}] \subset [\ddot{\omega},\ddot{\tau}]$,

$$\sqrt{n}[\,Vec(\widehat{\ddot{RM}}) - Vec(\ddot{RM})](t) \rightsquigarrow Z_{\ddot{RM}} \tag{2.16}$$

$$\sqrt{n}[\,Vec(\widehat{\ddot{RN}}) - Vec(\ddot{RN})](t) \rightsquigarrow Z_{\ddot{RN}}, \tag{2.17}$$

where $Z_{\ddot{RM}}$ *and* $Z_{\ddot{RN}}$ *are* p^2*-dimensional mean-zero Gaussian processes.*

The proof for Theorem 2 is provided in the Appendix, and uses similar methods to those of the proof for Theorem 1.

3. PCA Methods for Right-Censored Data

A PCA transforms a set of variables into linearly uncorrelated variables. For a $p \times 1$-dimensional random vector X, the first principal component direction, v_1, is the $p \times 1$ vector for which $\mathrm{Var}(v_1^T X)$ is maximized, subject to the constraint $||v_1||_2 = 1$. The jth principal component direction is the $p \times 1$ vector for which $\mathrm{Var}(v_j^T X)$ is maximized, subject to the constraints $||v_j||_2 = 1$ and $v_j^T v_{j'} = 0$, for $j' < j$. The principal components can be shown to be the eigenvectors of the covariance matrix of X. The solutions are not unique if there are repeated

eigenvalues. The principal components, $v_j^T X$, are linearly uncorrelated. The proportion of the variance of the data explained by the jth principal component is equal to $\lambda_j / \sum_{i=1}^{p} \lambda_i$, where λ_i is the ith eigenvalue of the covariance matrix of X. The PCA estimates are found using the eigendecomposition of the sample covariance matrix or the SVD of the mean-centered data.

In the presence of right censoring, it is not possible to nonparametrically estimate the principal components of T, the $p \times 1$-dimensional vector of failure times. This is because it is not possible to estimate the covariance matrix using standard methods without making strong assumptions on the form of the joint distribution. This is true both with and without the presence of competing risks. When there are no competing risks, instead of estimating the principal components for T, we consider the principal components for $N(\mathbf{t}) = [N^{(1)}(t_1), \ldots, N^{(p)}(t_p)]^T$ or $M(\mathbf{t}) = [M^{(1)}(t_1), \ldots, M^{(p)}(t_p)]^T$. The principal directions of $N(\mathbf{t})$ are the eigenvectors of $CN(\mathbf{t})$, and the principal directions of $M(\mathbf{t})$ are the eigenvectors of $CM(\mathbf{t})$. Similarly, in the presence of competing risks, we consider $\ddot{N}(\mathbf{t}) = [\ddot{N}_1^{(1)}(t_1), \ldots, \ddot{N}_1^{(p)}(t_p)]^T$, which has principal directions equal to the eigenvectors of $\dot{C\ddot{N}}(\mathbf{t})$, and $\ddot{M}(\mathbf{t}) = [\ddot{M}_1^{(1)}(t_1), \ldots, \ddot{M}_1^{(p)}(t_p)]^T$, which has principal directions equal to the eigenvectors of $\dot{C\ddot{M}}(\mathbf{t})$. In all cases, the correlation matrix can be used instead of the covariance matrix to obtain a scaled version of the PCA.

We obtain estimates for the principal directions of $N(\mathbf{t}), M(\mathbf{t}), \ddot{N}(\mathbf{t})$, and $\ddot{M}(\mathbf{t})$ using the eigenvectors of the consistent estimates of the relevant covariance or correlation matrices, which we derived in Section 2. Consistent estimates of the proportion of the variance explained are based on the corresponding eigenvalues. If we assume that the eigenvalues of $CM(\mathbf{t}), CN(\mathbf{t}), \dot{C\ddot{M}}(\mathbf{t})$, and $\ddot{CN}(\mathbf{t})$ are unique, then the estimates of the corresponding principal components are consistent and asymptotically normal, based on the results from Theorem 3. This is also true if we use the correlation matrices instead of the covariance matrices.

Theorem 3. *Assume that $t_1, \ldots, t_n$ are independent and identically distributed (i.i.d.) realizations of the $p \times 1$ random vector T, with joint distribution F_T. Assume that Σ is a $p \times p$ positive definite function of F_T, and that $\widehat{\Sigma}_n$ is a positive definite function of $(t_1, \ldots, t_n)$. Define $v_{\Sigma i}$ ($\hat{v}_{\hat{\Sigma} in}$) as the ith eigenvector of Σ ($\hat{\Sigma}$), and $\xi_{\Sigma i}$ ($\hat{\xi}_{\hat{\Sigma} in}$) as the ith eigenvalue of Σ ($\hat{\Sigma}$). Assume that $\sqrt{n}[Vec(\widehat{\Sigma}_n) - Vec(\Sigma)] \to_d N(0, \Psi_\Sigma)$, where Ψ_Σ is some positive definite matrix. Let $\Xi = diag(\xi_{\Sigma 1}, \ldots, \xi_{\Sigma p})$, $\hat{\Xi}_n = diag(\hat{\xi}_{\Sigma 1n}, \ldots, \hat{\xi}_{\Sigma pn})$, $V = [v_{\Sigma 1}, \ldots, v_{\Sigma p}]$, and $\hat{V} = [\hat{v}_{\hat{\Sigma} 1n}, \ldots, \hat{v}_{\hat{\Sigma} pn}]$. Then, $\sqrt{n}[vec(\hat{V}) - vec(V)] \to_d N(0, \Psi_V)$ and $\sqrt{n}[vec(\hat{\Xi}) -$*

$vec(\Xi)] \to_d N(0, \Psi_\Xi)$, where Ψ_V and Ψ_Ξ are positive semidefinite matrices.

The proof for Theorem 3 can be found in the Appendix, and follows similar steps to the proof in Anderson (2003) for a PCA using standard estimation techniques when the data have a multivariate normal distribution. When combined with the results from Section 2, this shows that the estimates for the principal component vectors based on $\widetilde{CM}(\mathbf{t}), \widetilde{CN}(\mathbf{t}), \widetilde{C\ddot{M}}(\mathbf{t})$, and $\widetilde{C\ddot{N}}(\mathbf{t})$, and the corresponding correlation matrix estimates are consistent and asymptotically normal.

The principal component scores can also be estimated for those subjects who have not been censored by time point $\mathbf{t}$. For this be the case, it must be that either $C > t_j$ or $\eta^{(j)} = 1$, for $j = 1, \ldots, p$. In this case we observe $N(\mathbf{t}) = [N^{(1)}(t_1), \ldots, N^{(p)}(t_p)]^T$, the entire vector of failure counting processes at time point $\mathbf{t}$. Define $\hat{v}_{cnj}(\mathbf{t})$ as the jth eigenvector of $\widehat{CN}(\mathbf{t})$. Then, the estimate of the jth principal component score is $\hat{v}_{cnj}(\mathbf{t})^T N(\mathbf{t})$. In the presence of a competing risk, a similar estimate can be made using $\ddot{N}_1(\mathbf{t}) = [\ddot{N}_1^{(1)}(t_1), \ldots, \ddot{N}_1^{(p)}(t_p)]^T$ and the jth eigenvector of $\widehat{C\ddot{N}}(\mathbf{t})$, $\hat{v}_{c\ddot{n}j}$. When a subject is not censored by time point $\mathbf{t}$, the entire vector of martingales, $M(\mathbf{t}) = [M^{(1)}(t_1), \ldots, M^{(p)}(t_p)]^T$, can be estimated consistently by plugging in a consistent estimate of $\Lambda^{(j)}(t_j \wedge T^{(j)})$, for $j = 1, \ldots, p$. If $\hat{v}_{cmj}(\mathbf{t})$ is the jth eigenvector of $\widehat{CM}(\mathbf{t})$, then the estimate of the jth principal component score is $\hat{v}_{cmj}(\mathbf{t})^T \hat{M}(\mathbf{t})$. Similar calculations can be performed in the case of a competing risk. In this case, the full vector of martingales, $\ddot{M}_1(\mathbf{t}) = [\ddot{M}_1^{(1)}(t_1), \ldots, \ddot{M}_1^{(p)}(t_p)]^T$, can be estimated by plugging in consistent estimates of $\ddot{\Lambda}_1^{(j)}(t_j \wedge T^{(j)})$, for $j = 1, \ldots, p$. The jth eigenvector of $\widehat{C\ddot{M}}(\mathbf{t})$ is defined as $\hat{v}_{c\ddot{m}_j}(\mathbf{t})$, and the estimate of the jth principal component score is $\hat{v}_{\widehat{cm}j}(\mathbf{t})^T \hat{M}_1(\mathbf{t})$. All of the estimates for the principal component scores based on the scaled counting processes or martingales can be estimated in the same way using the eigenvectors of the correlation matrices. In Section 5, we provide an example in which the principal component score estimates are used as covariates in a Cox PH model. This is possible because the form of the partial likelihood for the Cox PH model requires only the covariate values for those subjects who have not been censored by a given timepoint. This allows us to use principal component scores for censored time-to-event variables as covariates in a Cox PH model, similarly to how principal component scores are used for a principal component regression.

The estimation techniques for the PCA derived above all allow for right censoring from loss to follow up or administrative censoring after reaching the end of the follow-up period. For certain censoring schemes, there may be simpler ways to estimate the principal components for $M(\mathbf{t})$ and $N(\mathbf{t})$. However, we

believe that for almost all types of time-to-event data, it will be necessary to estimate the principal components of $M(\mathbf{t})$ and $N(\mathbf{t})$, rather than T directly. Estimating the principal components for T would require no censoring, implying an unlimited follow-up time and no competing risks, both of which are rare for data sets and studies with time-to-event data. However, one setting in which we can simplify the estimation of the covariance or correlation of $M(\mathbf{t})$ and $N(\mathbf{t})$ is that in which there is only administrative censoring and no loss to follow up. In this case, we assume that all subjects have the same censoring time, $C = c$. Note that unless all subjects have had all of the event types by time c, it is still not possible to estimate the principal components for T. However, the estimation of the principal components for $M(\mathbf{t})$ or $N(\mathbf{t})$ can be simplified for $t_1, \ldots, t_p < c$. In this case, $N(\mathbf{t})$ is fully observed for each subject, because none have been censored before the specified timepoints. Thus, the principal components can be estimated using a standard estimation of the covariance matrix for $N(\mathbf{t})$. Similarly, $M(\mathbf{t})$ involves an indicator variable that is observed for all subjects, and a set of cumulative hazard functions that can all be estimated up to time c. This means we can obtain an estimate of $M(\mathbf{t})$ for each subject, and again use standard covariance and correlation estimation techniques to obtain the principal component estimates for $M(\mathbf{t})$. Similarly, in the competing-risk setting, if there is no loss to follow up and only administrative censoring at time c, it is still not possible to estimate the principal components of T, unless every subject has had each of the events before the censoring time, c, and before they have the competing event. However, in this set up, the necessary indicator functions $\ddot{N}_l^{(j)}(t)$ are observed for all event types up to time c, and the necessary cause-specific hazard functions can also be estimated up to time c. Therefore, as before, we observe $\ddot{N}(\mathbf{t})$ and can obtain estimates of $\ddot{M}(\mathbf{t})$ for each subject $t_1, \ldots, t_p < c$. From here, standard covariance and correlation estimation techniques can be used to estimate the principal components for $\ddot{N}(\mathbf{t})$ and $\ddot{M}(\mathbf{t})$.

A question of interest for researchers is whether a PCA based on counting processes or martingales is preferred in practice. Although each has its advantages, we believe that in general, a PCA based on martingales is more useful, because it contains more information. This is because, as noted previously, $M^{(j)})(t)$ and $\ddot{M}^{(j)}(t)$ contain information on whether event type j has happened by timepoint t and when it happened, assuming it happened before timepoint t. In contrast, $N^{(j)})(t)$ and $\ddot{N}^{(j)}(t)$ contain information only on whether event type j has happened by timepoint t.

4. Simulation Results

Simulations were conducted to examine the estimations of the principal components of $M(\mathbf{t}), N(\mathbf{t})$, $\ddot{M}_1(\mathbf{t})$, and $\ddot{N}_1(\mathbf{t})$, focusing on the results using the correlation matrices. Data sets were simulated with $p = 8, 16$, and 32. First, we simulate $W = [W^{(1)}, \ldots, W^{(p)}]$ using a multivariate normal distribution, where $W^{(j)}$ has mean zero and standard deviation one, for $j = 1, \ldots, p$. For each of the dimensions, the covariance matrix of W is a block-diagonal matrix, with four equal size blocks along the diagonals, of the form

$$\Sigma_W = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_4 \end{bmatrix}.$$

Here, $\mathbf{A}_1$ has all ones along the diagonal and 0.7 on all off-diagonal elements, $\mathbf{A}_2$ has all ones along the diagonals and 0.4 on all off-diagonal elements, $\mathbf{A}_3$ has all ones along the diagonal and 0.2 on all off-diagonal elements, and $\mathbf{A}_4$ has all ones along the diagonal and 0.1 along all off-diagonal elements. Furthermore, $T = [T^{(1)}, \ldots, T^{(p)}]$ is defined as a transformation of W such that $T^{(j)} \sim$ Exponential(1). Specifically, $T^{(j)} = -\ln\{1 - \Phi(W^{(j)})\}$, where $\Phi(\cdot)$ is the CDF of a standard normal distribution.

The competing risk setting has a similar setup: W^T has the same distribution as above, $[W^T, \ddot{W}]^T$ has a multivariate normal distribution, and the correlation between $\ddot{W}$ and $W^{(j)}$ is 0.1, for $j = 1, \ldots, p$. Furthermore, T is still defined as the same transformation of W and $\ddot{T} = -\ln\{1 - \Phi(\ddot{W})\}$.

In all settings, the censoring distribution is $[1/4 \cdot C_1] \sim$ Beta$(1.5, 6.5)$. Simulations with the censoring distribution $C_2 \sim$ Uniform$(0, 4)$ are provided in the Supplementary Material. For all settings, sample sizes of $n = 200$ and $n = 1000$ are simulated. In the setting without competing risks, the average censoring rate for each event is 53% when using C_1, regardless of p, and 25% when using C_2. In the semi-competing risk setting, the censoring rate is 34% when using C_1, and 13% when using C_2. For both censoring schemes in the competing risk setting, the competing and noncompeting events are equally likely to be observed. The lower censoring rate in the competing risk setting is due to the fact that the censoring time must come before both the noncompeting event time and the competing event time in order for the subject to be censored.

For each data set, we estimated $\widetilde{RN}(\mathbf{t}), \widetilde{RM}(\mathbf{t}), \widetilde{R\ddot{N}}(\mathbf{t})$, and $\widetilde{R\ddot{M}}(\mathbf{t})$ using the Dabrowska estimator for bivariate survival functions, the Kaplan–Meier

estimator for univariate survival functions, and the Nelson–Aalen estimator for all cumulative hazard functions. In order to ensure that all estimated covariance and correlation matrices are positive definite we use a minimum eigenvalue of 0.001. Matrices are estimated at $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$ and $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$. The true covariance matrices are calculated empirically by simulating a single data set with 500,000 subjects and no censoring, which allows us to use standard correlation and covariance estimation methods. In the competing risk setting, the cumulative hazard necessary to calculate the martingales is estimated using the Nelson–Aalen estimator based on the 500,000 simulated subjects. In the setting without competing risks, the cumulative hazard is known, based on the distribution of $T^{(j)}$. The true principal component directions are calculated as the eigenvectors of the true correlation and covariance matrices. More information on the true correlation matrices can be found in the Supplementary Material.

For the ith simulated data and jth principal component direction, the angle, in radians, between the estimated direction, $\hat{v}_{ij}$, and the true direction, v_j, is calculated as $\text{Angle}(\hat{v}_{ij}, v_j) = \cos^{-1}[|\hat{v}_{ij}^T v_j|/(||\hat{v}_{ij}||_2 \cdot ||v_j||_2)]$. For each setting, the bias and standard deviation for the jth principal direction are calculated as the empirical mean and standard deviation, respectively, of the angle between the estimated direction and the true direction for all 1,000 simulated data sets. Table 1 reports the average angles and standard deviations for the principal component directions using the eigendecompositions of $\widetilde{RN}(\mathbf{t})$, $\widetilde{RM}(\mathbf{t})$, $\ddot{\widetilde{RN}}(\mathbf{t})$, and $\ddot{\widetilde{RM}}(\mathbf{t})$ for C_1 and $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$. The results for $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ and C_2 are reported in the Supplementary Material. In all cases, just the first four directions are reported, owing to the four-block structure of the underlying covariance and correlation matrices. Table 1 shows that, as expected, the average angle decreases as the sample size increases. Furthermore, the average angle for the first direction actually decreases as the number of dimensions increases. This is likely due to the increase in the leading eigenvalue for higher dimension setups. In our higher dimension setups the leading principal component explains a larger percentage of the overall variance than it does in lower dimension setups. In addition, because of the nature of our block matrix setup, as the dimension of a setup increases the difference between the first and second eigenvalues also increases. Both of these facts likely contribute to the first eigenvector being estimated more precisely in our simulations. The true eigenvectors and eigenvalues of the matrices can be found in the Supplementary Material. For $RM(\mathbf{t})$, when $n = 1000$, the average angle between the estimated first direction and true first direction is less than 0.30 radians, which is just over 17 degrees. Even 0.58 radians, which is the highest average angle between the true and the estimated first principal directions for

Table 1. Average (SD) of the angle, in radians, between the true and estimated PCA directions based on counting process and martingale correlations using the censoring distribution C_1.

		t=1					
		No competing risks					
		Counting Process			Martingale		
		n=200	n=1000	Eigenvalue	n=200	n=1000	Eigenvalue
8 Dim	PC 1	0.70 (0.24)	0.35 (0.12)	1.49	0.58 (0.19)	0.28 (0.09)	1.59
	PC 2	0.98 (0.30)	0.56 (0.22)	1.26	0.84 (0.29)	0.43 (0.16)	1.32
	PC 3	1.12 (0.29)	0.78 (0.32)	1.13	1.00 (0.32)	0.62 (0.28)	1.16
	PC 4	1.15 (0.29)	0.82 (0.35)	1.06	1.04 (0.32)	0.62 (0.32)	1.08
16 Dim	PC 1	0.46 (0.14)	0.21 (0.06)	2.47	0.37 (0.10)	0.17 (0.05)	2.77
	PC 2	0.73 (0.25)	0.32 (0.09)	1.77	0.54 (0.18)	0.24 (0.07)	1.96
	PC 3	1.03 (0.28)	0.49 (0.19)	1.37	0.82 (0.28)	0.35 (0.12)	1.47
	PC 4	1.19 (0.25)	0.65 (0.26)	1.19	1.04 (0.30)	0.42 (0.15)	1.24
32 Dim	PC 1	0.37 (0.08)	0.17 (0.04)	4.41	0.30 (0.07)	0.14 (0.04)	5.13
	PC 2	0.55 (0.15)	0.24 (0.05)	2.79	0.42 (0.10)	0.19 (0.04)	3.25
	PC 3	0.86 (0.25)	0.35 (0.07)	1.87	0.62 (0.19)	0.26 (0.05)	2.11
	PC 4	1.17 (0.24)	0.50 (0.11)	1.43	0.92 (0.26)	0.34 (0.06)	1.55
		With competing risk					
8 Dim	PC 1	0.41 (0.28)	0.15 (0.06)	2.20	0.66 (0.23)	0.35 (0.12)	1.56
	PC 2	0.95 (0.38)	0.61 (0.40)	1.19	0.93 (0.30)	0.55 (0.23)	1.30
	PC 3	1.15 (0.29)	0.91 (0.38)	1.02	1.07 (0.31)	0.74 (0.30)	1.15
	PC 4	1.17 (0.28)	0.94 (0.37)	0.93	1.14 (0.27)	0.80 (0.34)	1.07
16 Dim	PC 1	0.36 (0.23)	0.14 (0.04)	3.78	0.45 (0.14)	0.21 (0.07)	2.69
	PC 2	0.83 (0.38)	0.40 (0.28)	1.85	0.68 (0.22)	0.31 (0.10)	1.90
	PC 3	1.09 (0.31)	0.69 (0.35)	1.34	0.94 (0.28)	0.47 (0.18)	1.43
	PC 4	1.16 (0.26)	0.77 (0.31)	1.08	1.17 (0.26)	0.62 (0.23)	1.21
32 Dim	PC 1	0.35 (0.22)	0.13 (0.03)	6.99	0.37 (0.09)	0.17 (0.04)	4.93
	PC 2	0.74 (0.36)	0.31 (0.19)	3.14	0.54 (0.15)	0.24 (0.06)	3.10
	PC 3	0.98 (0.30)	0.54 (0.27)	1.98	0.80 (0.24)	0.36 (0.10)	2.01
	PC 4	1.02 (0.24)	0.64 (0.21)	1.36	1.07 (0.24)	0.51 (0.11)	1.49

$RM(\mathbf{t})$ at $n = 200$, is just over 33 degrees. At $\mathbf{t} = [1, 1, 1, 1, \ldots]^T$, the estimates based on $RM(\mathbf{t})$ outperform those based on $RN(\mathbf{t})$ across all sample sizes and dimensions. This is also true when $\mathbf{t} = [2, 2, 2, 2, \ldots]^T$ and C_2, as shown in the Supplementary Material. A key reason for this is that the variance of the counting processes is much lower than the variances of the martingales, leading to less precise estimates of the relevant correlations.

In the competing risk setting, the angle between the true and estimated principal directions based on $\ddot{RM}(\mathbf{t})$ are similar, but slightly worse than those based on $RM(\mathbf{t})$ in the noncompeting risk section. However, in the competing risk setting, the average angle between the estimated and true principal directions using $\ddot{RN}(\mathbf{t})$ are lower than those using $\ddot{RM}(\mathbf{t})$. This is again likely due to the

leading eigenvalues for $\ddot{R}\ddot{N}(\mathbf{t})$ being larger, with more separation than $RN(\mathbf{t})$. This is because $\ddot{N}_1^{(j)}(t)$ and $\ddot{N}_1^{(j')}(t)$ tend to be correlated, even when $T^{(j)}$ and $T^{(j')}$ are uncorrelated. Consider the fact that in order for $\ddot{N}_1^{(j)}(t)$ to equal one, it must be that both $T^{(j)} \leq t$ and $T^{(j)} \leq \ddot{T}$. If $T^{(j)}$ and $T^{(j')}$ are independent and uncorrelated, $P(T^{(j')} \leq \ddot{T} | T^{(j)} \leq \ddot{T})$ is higher than $P(T^{(j')} \leq \ddot{T})$. This, in turn, means that when $T^{(j)}$ and $T^{(j')}$ are independent, $P(\ddot{N}_1^{(j')}(t) = 1 | P(\ddot{N}_1^{(j)}(t) = 1)$ is larger than $P(\ddot{N}_1^{(j')}(t) = 1)$, leading to a positive correlation between $\ddot{N}_1^{(j)}(t)$ and $\ddot{N}_1^{(j')}(t)$. The effect on $\ddot{M}_1^{(j)}(t)$ and $\ddot{M}_1^{(j')}(t)$ is not as large, because it does not introduce correlation between $\ddot{\Lambda}_1^{(j)}(t \wedge \ddot{T}^{(j)})$ and $\ddot{\Lambda}_1^{(j')}(t \wedge \ddot{T}^{(j')})$ to the same degree. Because of this, in our simulation setup the separation between the first and second eigenvalues for $\ddot{R}\ddot{N}(\mathbf{t})$ is larger than that for $\ddot{R}\ddot{M}(\mathbf{t})$, which leads to more precise estimates of the first principal direction. More information on this, and specifically the correlation between $\ddot{N}^{(j)}(t)$ and $\ddot{N}^{(j')}(t)$, even when $T^{(j)}$ and $T^{(j')}$ are uncorrelated, can be found in the Supplementary Material.

5. MPACT Trial

The MPACT trial was a clinical trial that ran from 2009 to 2013, in which 861 patients with metastatic pancreatic cancer were randomized to be treated with either the standard gemcitabine, or a novel medication called paclitaxel (Von Hoff et al. (2013)). The data for the 430 patients randomized to the standard care are available through Project Data Sphere®.

During this trial, nine adverse events occurred in at least 50 of the patients: abdominal pain, anemia, constipation, decreased appetite, fatigue, nausea, neutropenia, thrombocytopenia, and vomiting. For each adverse event, the failure time is the time from randomization to the first occurrence of that event. Patients who left the study as a result of death or disease progression before having a given adverse event are considered to have a competing event. Patients who left the study before having a given adverse event for any other reason are considered to be censored. Table 2 gives an overview of the nine events. The event rate is the proportion of subjects who had the given adverse event, and the censoring rate is the proportion of subjects who were censored. The median and mean event times are calculated only among those who had the given event of interest. We can see that the event rate ranges from 0.13 to 0.27, whereas the censoring rate ranges from 0.28 to 0.35. In addition, the distribution of all nine event types is positively skewed with the mean larger than the median, indicating that most subjects who have a given event have it early in the study, with a few having the event later in the study.

Table 2. Summary statistics for the nine adverse events considered from the MPACT trial.

	Event Rate	Censoring Rate	Median Time (Days)	Mean Time (Days)
Abdominal Pain	0.14	0.34	36	57.31
Anemia	0.26	0.29	30	59.10
Constipation	0.13	0.35	18.5	43.09
Decreased Appetite	0.14	0.33	22.5	46.50
Fatigue	0.27	0.29	19	39.96
Nausea	0.20	0.33	15	35.86
Neutropenia	0.26	0.28	23	53.43
Thrombocytopenia	0.18	0.31	19	69.68
Vomiting	0.14	0.35	24	41.95

In order to better understand the relationship between the nine adverse events, we conduct a survival PCA for these events. We use the martingale correlation matrix, because the martingales contain information on when an event occurred and whether it occurred by a given time. Because patients are subject to both censoring and the competing risk of death or the progression of the disease, we use the competing risk martingale correlation matrix. Here $\ddot{RM}(\mathbf{t})$ is estimated for timepoints between $\mathbf{t} = [30, 30, \ldots]^T$ and $\mathbf{t} = [360, 360, \ldots]^T$, in increments of one day. Day 360 was chosen as the final timepoint because by that time, 420 of the 430 patients had left the study. Day 30 was chosen as the starting point to ensure that a sufficient number of events had occurred in order to get reasonable precision for the estimates. The principal component loadings based on $\ddot{RM}(\mathbf{t})$ at $\mathbf{t} = [360, 360, \ldots]^T$ are presented in Table 3. Figure 1 shows the directions for the first two principal components, plotted over time between day 30 and 360. The line type is based on the following clinically defined groupings: constitutional (C): fatigue; gastrointestinal (G): abdominal pain, constipation, decreased appetite, nausea, vomiting; hematologic (H): anemia, neutropenia, thrombocytopenia.

Figure 1 shows that the largest loadings in the first principal component are gastrointestinal and constitutional events, and the loadings for these events all go in the same direction. The largest loadings for the second principal component are the hematologic events, which also all go in the same direction. Therefore, the first principal component is driven by the occurrence of gastrointestinal events and fatigue, and the second principal component is driven by the occurrence of hematologic events. This shows that the martingales for the gastrointestinal and constitutional events tend to be correlated with the martingales for other gastrointestinal and constitutional events, whereas the martingales for the hematologic

Table 3. Principal component directions at day 360 and the proportion of the variance explained for each principal component using estimates based on a martingale correlation matrix.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Abdominal pain	0.38	-0.21	-0.18	0.03	0.05	0.73	-0.41	-0.26	-0.04
Anemia	0.09	0.42	0.56	0.51	-0.39	0.26	-0.02	0.15	-0.05
Constipation	0.25	0.26	0.36	-0.16	0.80	0.10	0.09	0.21	-0.11
Decreased appetite	0.42	-0.16	-0.42	0.33	-0.03	-0.09	0.11	0.68	-0.16
Fatigue	0.43	0.00	-0.06	0.47	0.17	-0.28	0.28	-0.53	0.34
Nausea	0.46	0.14	0.06	-0.35	-0.29	-0.21	0.12	-0.27	-0.65
Neutropenia	-0.10	0.55	-0.45	-0.14	-0.06	0.40	0.54	-0.04	0.09
Thrombocytopenia	-0.01	0.60	-0.34	0.09	0.10	-0.30	-0.65	-0.04	-0.01
Vomiting	0.44	0.10	0.14	-0.48	-0.27	-0.07	-0.09	0.22	0.64
Proportion Variance	0.26	0.15	0.11	0.10	0.10	0.08	0.08	0.06	0.05

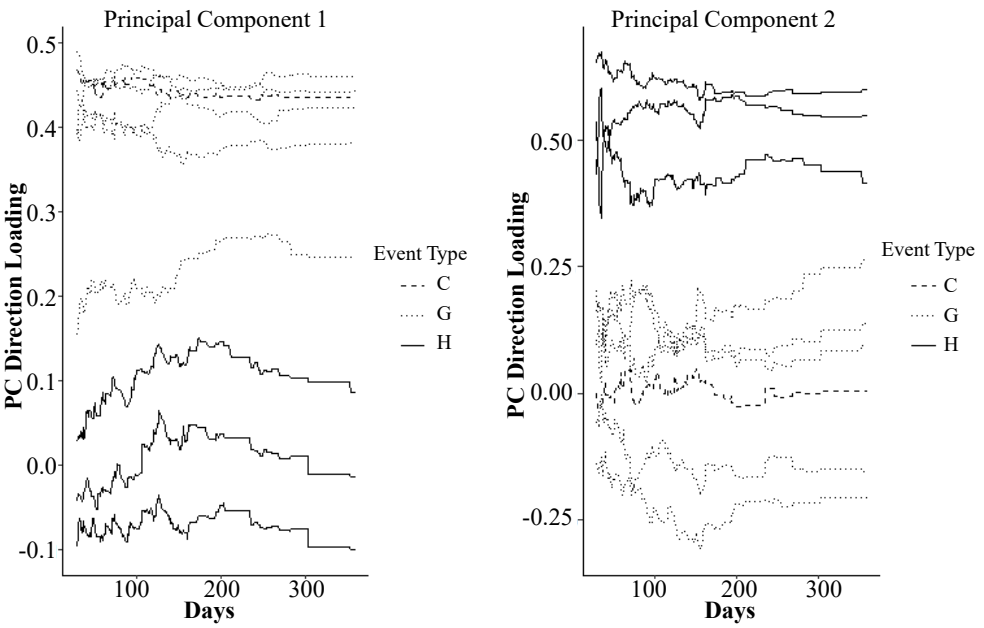

Figure 1. Principal component direction loadings from day 30 to 360 for the first two principal components using martingale correlation matrix estimates. Line types indicate constitutional, gastrointestinal, and hematologic event types.

events tend to be correlated with those for other hematologic events. Together, the first two principal components explain close to 40% of the total variance, which is consistent over time.

Unlike the first two principal components, the third through ninth principal components do not have a straightforward interpretation. In addition, as shown in Table 3, the proportions of the variance explained for the third through seventh principal components are similar. This is true for all the time points estimated, and caused potentially crossing eigenvalues. In order to make them comparable over time, instead of ordering the principal directions using the proportion of the variance explained, the order was chosen to minimize the sum of the angles between the principal directions across time. First, day 360 was chosen as a reference date, and the principal directions were ordered in descending order of the associated eigenvalues. We define $\hat{v}_{jt}$ as the jth principal direction for $\mathbf{t} = [t, t, \ldots]^T$. For all days other than 360, the ordering of the principal directions was chosen to minimize $\sum_{j=1}^{9} \text{Angle}(\hat{v}_{jt}, \hat{v}_{j360})$. This meant that for some days, the ordering of the principal directions changed. A figure of the principal directions using this method can be found in the Supplementary Material.

In addition to the simple analysis of the principal component loadings, we used the first two principal component scores in a Cox PH model, with death or progression of the disease as the outcome. We follow the methods described in Section 3 to estimate the principal component scores. Because the loadings and interpretation of the first two principal components are consistent over time, we include the principal component scores as time-varying covariates. The first principal component can be thought of as a measure of how often and how early a subject had gastrointestinal and constitutional events, with those subjects who have had more and earlier such events tending to have a higher score. Similarly, the second principal component can be thought of as a measure of how often and how early a subject had hematologic events. The scores for the first two principal components were then used as time-varying covariates in a Cox PH model, with time until death or progression of the disease as the outcome of interest. In addition to the principal component scores, the age in years, sex, and Karnofsky performance status were included as covariates. The Karnofsky performance status is a numeric measure of the general wellbeing of cancer patients, and ranges from 70 to 100 within the sample (Karnofsky et al. (1948)). One subject was dropped from the model because of a missing Karnofsky performance status.

Table 4 gives the estimated hazard ratios for the Cox PH model. The hazard ratio for the first principal component is estimated to be 1.62, and has a p-value of 4.28×10^{-7}. This indicates that those subjects who have had more and earlier gastrointestinal events by a given timepoint have a higher estimated hazard of death, holding the other covariates constant. This result is consistent with the notion that as-yet-undocumented disease progression may manifest as

Table 4. Estimated hazard ratios for Cox PH models, including the first two martingale correlation PC scores.

	HR	P-Value
PC1	1.62	4.28E-07
PC2	0.92	0.46
Age	0.99	0.04
Sex: Male	1.01	0.94
Karnofsky Performance Status	0.98	0.02
N	429	
Events	264	

fatigue or gastrointestinal adverse events. This means that subjects with a higher PC 1 score have a higher hazard for death or progression of the disease, as we see in the results. The hazard ratio for the second principal component is less than one, but the p-value is equal to 0.46. The estimated coefficient for PC 2 indicates that those subjects who have had more and earlier hematologic events by a given time point have a lower hazard of death, holding other covariates constant. However, given the p-value, we cannot rule out a null or opposite result. PC 2 having a hazard ratio below one is consistent with hematologic adverse events indicating greater exposure to the drug (owing to increased absorption or decreased degradation) or a greater sensitivity to its effects. In this case, we might expect subjects with more and earlier hematologic events to have a lower hazard of death or progression of the disease. This is consistent with the estimated hazard ratio for PC 2 being below one. However the p-value indicates that this may just be due to random variation in the MPACT trial sample. As a sensitivity analysis, we estimated the principal components using the counting process correlation matrix instead of the martingale correlation matrix, and estimated the same Cox PH model with the counting process correlation principal component scores. These results led to similar conclusions, and can be found in the Supplementary Material.

6. Conclusion

We have shown how a PCA can be defined for multivariate survival data in the presence of censoring by using either the counting processes or the corresponding martingales defined by each event type. We build on previous results for bivariate survival data to show how to estimate the full covariance and correlation matrices for either the counting processes or the martingales at a given time point. In addition, we extended this to the semicompeting-risk setting, in which each of

the event types is subject to a competing risk and an independent censoring time. For both the standard censoring-only setting and the semicompeting-risk setting, we show that the estimators for the martingale and counting process covariance or correlation matrices converge to a mean-zero Gaussian process when properly normalized.

We also show that the loadings for the principal components based on the martingales and counting processes can be estimated consistently using the eigendecomposition of the corresponding covariance or correlation matrix. The corresponding estimates are shown to be consistent and asymptotically normal. A subject's principal component score can only be estimated up to the time they are censored. However, this still allows for principal component scores to be used as covariates in a Cox PH regression.

Our results are all for a fixed number of noncompeting failure times, p. Given the increasing importance of "big data," it is also of interest to consider the high-dimensional setting where $p \to \infty$. Here, it would be important to define a set of conditions under which the estimates of the martingale and counting process covariance and correlation matrices converge in probability to the corresponding true matrices. One way to do this is to use maximal inequalities, as defined in Kosorok (2008). This would allow us to show convergence, even as $p \to \infty$, as long as n grows at a fast enough rate relative to p. However, in order to use maximal inequalities, it is necessary to sufficiently bound the rate at which the martingale or counting process covariance and correlation estimates converge to the true value for each pair of failure times. This is a stronger condition than the consistency and asymptotic normality shown here. An example of the type of bound needed can be found in Theorem 1 of Bitouzé, Laurent and Massart (1999), which defines a Dvoretzky–Kiefer–Wolfowitz-type inequality for the Kaplan–Meier estimator. To the best of our knowledge, no similar result has been shown for any known bivariate survival function estimator. If such a result can be shown to hold, then the consistency of the martingale and counting process correlation and covariance estimates in a high-dimensional setting would follow.

The usefulness of this method is shown using data from the comparator arm of the MPACT trial for patients with pancreatic cancer. We estimate the principal components based on the martingale correlation matrix for nine adverse events experienced by patients in the trial, and define medically relevant groupings of these events. An R package that implements these methods is available at `https://github.com/blangworthy/survPCA`. One area for future research is to further consider ordering and potential changes to the principal components over time. When two or more principal components have similar eigenvalues, the

ordering may change over time, owing to random noise. It may be of interest to investigate how to identify when changes in principal component loadings and eigenvalues over time are due to random noise or to a change in the true underlying covariance or correlation matrix. In addition, the described methods do not allow for us to control for covariates. For a standard PCA, it is possible to estimate a conditional PCA by using model residuals after regressing out a set of shared covariates. A similar idea may be possible using a Cox PH model for each event type using a shared set of covariates, and then estimating the PCA using the martingale residuals. One difference between this method and the survival PCA presented here is that the martingale residuals are typically calculated at the event or censoring time for each subject, which can vary between subjects and event types in a multivariate setting. However, martingales always have a zero mean. Therefore, the covariance measures between them may give a meaningful measure of the association between different event types.

Supplementary Material

The online Supplementary Material includes additional simulation results and an additional analysis of the data from the MPACT trial.

Acknowledgments

The authors thank Dr. Yu Cheng from the University of Pittsburgh for supplying the code to compute the Dabrowska bivariate survival function estimator.

Appendix

Proof of Theorem 1. Define $\theta_C = [S^{(1)}, \ldots, S^{(p)}, \Lambda^{(1)}, \ldots, \Lambda^{(p)}, S^{(1,2)}, \ldots, S^{(p-1,p)}]^T$ and $\hat{\theta}_C = [\hat{S}^{(1)}, \ldots, \hat{S}^{(p)}, \hat{\Lambda}^{(1)}, \ldots, \hat{\Lambda}^{(p)}, \hat{S}^{(1,2)}, \ldots, \hat{S}^{(p-1,p)}]^T$ where $\hat{S}^{(j)}$ is the Kaplan–Meier estimator, $\hat{\Lambda}^{(j)}$ is the Nelson–Aalen estimator. We will assume $\hat{S}^{(j,j')}$ is the Dabrowska estimator, but other bivariate survival estimators are possible. Based on previously shown results for the Kaplan–Meier, Nelson–Aalen, and Dabrowska estimators, (see Kosorok (2008) for further details on the Kaplan–Meier and Nelson–Aalen estimators and Cheng, Fine and Kosorok (2007) for further details on the Dabrowska estimator)

$$\sqrt{n}[\hat{\theta}_C - \theta_C] \to Z_{\theta_C},$$

where $Z_{\theta_C} = [Z_{S_1}, \ldots, Z_{S_p}, Z_{\Lambda_1}, \ldots, Z_{\Lambda_p}, Z_{S_{1,2}}, \ldots, Z_{S_{p-1,p}}]^T$ is a $2p + (p^2 - p)/2$ dimensional mean 0 Gaussian process. Using this set up we can show the results

for $\widehat{CN}$ and $\widehat{CM}$

1. $\widehat{CN}$: Consider the mapping

$$\phi_{CNjj'}(\theta_C) = S^{(j,j')} - S^{(j)}S^{(j')}.$$

Using this mapping

$$\sqrt{n}[\text{Vec}(\widehat{CN}) - \text{Vec}(CN)] = \sqrt{n}\begin{bmatrix} \phi_{CN11}(\hat{\theta}_C) - \phi_{CN11}(\theta_C) \\ \phi_{CN12}(\hat{\theta}_C) - \phi_{CN12}(\theta_C) \\ \vdots \\ \phi_{CNp-1p}(\hat{\theta}_C) - \phi_{CNp-1p}(\theta_C) \\ \phi_{CNpp}(\hat{\theta}_C) - \phi_{CNpp}(\theta_C) \end{bmatrix}.$$

The Hadamard derivative of $\phi_{CNjj'}(\theta_C)$ in the direction of Z_{θ_C} is

$$\phi'_{\theta_C CNjj'}(Z) = S^{(j)}Z_{S'_j} + S^{(j')}Z_{S_j} - Z_{S_{j,j'}}.$$

Therefore by the functional delta method, Theorem 2.8 in Kosorok (2008), $\sqrt{n}[\text{Vec}(\widehat{CN}) - \text{Vec}(CN)] \rightsquigarrow Z_{CN}$ where Z_{CN} is a p^2 dimensional mean 0 Gaussian process.

2. $\widehat{CM}$: Consider the mapping

$$\begin{aligned}\phi_{CMjj'}(\theta_C)(t_j, t_{j'}) = S^{(jj')}(t_j, t_{j'}) - 1 + \int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'})\lambda^{(j)}(s_j)ds_j + \\ \int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j'}^-)\lambda^{(j')}(s_{j'})ds_{j'} + \\ \int_0^{t_j}\int_0^{t_{j'}} S^{(j,j')}(s_j^-, s_{j'}^-)\lambda^{(j)}(s_j)\lambda^{(j')}(s_{j'})ds_j ds_{j'}. \quad \text{(A.1)}\end{aligned}$$

Using this mapping

$$\sqrt{n}[\text{Vec}(\widehat{CM}) - \text{Vec}(CM)] = \sqrt{n}\begin{bmatrix} \phi_{CM11}(\hat{\theta}_C) - \phi_{CM11}(\theta_C) \\ \phi_{CM12}(\hat{\theta}_C) - \phi_{CM12}(\theta_C) \\ \vdots \\ \phi_{CMp-1p}(\hat{\theta}_C) - \phi_{CMp-1p}(\theta_C) \\ \phi_{CMpp}(\hat{\theta}_C) - \phi_{CMpp}(\theta_C) \end{bmatrix}.$$

As with the results for $\widehat{CN}$, the desired result will follow if it can be shown that $\phi_{CMjj'}(\theta_C)(t_j, t_{j'})$ is Hadamard differentiable in the direction of Z_{θ_C}. In order to do this we can consider each of the five parts on the right

hand side of Equation (A.1) separately. It is straightforward to show that $S^{(j,j')}(t_j, t_{j'})$ and -1 are hadamard differentiable. $\int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'})\lambda^{(j)}(s_j)ds_j$ and $\int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j'}^-)\lambda^{(j')}(s_{j'})ds_{j'}$ can both be shown to be Hadamard differentiable through Lemma 12.3 from Kosorok (2008), and $\int_0^{t_j}\int_0^{t_{j'}} S^{(j,j')}(s_j^-, s_{j'}^-)$ $\lambda^{(j)}(s_j)\lambda^{(j')}(s_{j'})ds_j ds_{j'}$ can be shown to be Hadamard differentiable through Lemma A5 from Cheng, Fine and Kosorok (2007). Therefore by the functional delta method $\sqrt{n}[\text{Vec}(\widehat{CM}) - \text{Vec}(CM)] \rightsquigarrow Z_{CM}$ where Z_{CM} is a p^2 dimensional mean 0 Gaussian process.

In order to show the results for $\widehat{RN}$ and $\widehat{RM}$ we define $\theta_R = [S^{(1)}, \ldots, S^{(p)}, CN^{(1,1)}, CN^{(1,2)}, \ldots, CN^{(p,p)}, CM^{(1,1)}, CM^{(1,2)}, \ldots, CM^{(p,p)}]^T$ and $\hat{\theta}_R = [\hat{S}^{(1)}, \ldots, \hat{S}^{(p)}, \widehat{CN}^{(1,1)}, \widehat{CN}^{(1,2)}, \ldots, \widehat{CN}^{(p,p)}, \widehat{CM}^{(1,1)}, \widehat{CM}^{(1,2)}, \ldots, \widehat{CM}^{(p,p)}]^T$. Then using results from above

$$\sqrt{n}[\hat{\theta}_R - \theta_R] \rightsquigarrow Z_{\theta_R},$$

where $Z_{\theta_R} = [Z_{S_1}, \ldots, Z_{S_p}, Z_{CN_{1,1}}, Z_{CN_{1,2}}, \ldots, Z_{CN_{p,p}}, Z_{CM_{1,1}}, Z_{CM_{1,2}}, \ldots, Z_{CM_{p,p}}]^T$ is a $p + 2p^2$ dimensional mean 0 Gaussian process. From here we can show the results for $\widehat{RN}$ and $\widehat{RM}$ using the functional delta method as above.

1. $\widehat{RN}$ Define

$$\phi_{RNjj'}(\theta_R) = \frac{CN^{(j,j')}}{\sqrt{S^{(j)}(1 - S^{(j)})}\sqrt{S^{(j')}(1 - S^{(j')})}}.$$

Using this mapping

$$\sqrt{n}[\text{Vec}(\widehat{RN}) - \text{Vec}(RN)] = \sqrt{n}\begin{bmatrix} \phi_{RN11}(\hat{\theta}_R) - \phi_{RN11}(\theta_R) \\ \phi_{RN12}(\hat{\theta}_R) - \phi_{RN12}(\theta_R) \\ \vdots \\ \phi_{RNp-1p}(\hat{\theta}_R) - \phi_{RNp-1p}(\theta_R) \\ \phi_{RNpp}(\hat{\theta}_R) - \phi_{RNpp}(\theta_R) \end{bmatrix}.$$

The Hadamard differentiability of $\phi_{RNjj'}(\theta_R)$ in the direction of Z_{θ_R} can be shown through repeated application of the chain rule and the following results,

- If $\phi_1(A) = A^2$ then $\phi'_{A1}(\alpha) = 2A\alpha$.
- If $\phi_2(A) = \sqrt{A}$ then $\phi'_{A2}(\alpha) = \alpha/(2\sqrt{A})$.
- If $\phi_3(A, B) = A/B$ then $\phi'_{AB3}(\alpha, \beta) = (B\alpha - A\beta)/B^2$.

Therefore by the functional delta method $\sqrt{n}[\text{Vec}(\widehat{RN}) - \text{Vec}(RN)] \rightsquigarrow Z_{RN}$ where Z_{RN} is a p^2 dimensional mean 0 Gaussian process.

2. $\widehat{RM}$ Results are similar to those for $\widehat{RN}$ if we consider the mapping

$$\phi_{RMjj'}(\theta_R) = \frac{CM^{(j,j')}}{\sqrt{1-S^{(j)}}\sqrt{1-S^{(j')}}}.$$

Proof of Theorem 2. Similar to the proof of Theorem 1 we define $\ddot{\theta}_C = [F_1^{(1)}, \ldots, F_1^{(p)}, F_{11}^{(1,1)}, F_{11}^{(1,2)}, \ldots, F_{11}^{(p,p)}, F_{12}^{(1,1)}, F_{12}^{(1,2)}, \ldots, F_{12}^{(p,p)}, F_{21}^{(1,1)}, F_{21}^{(1,2)}, \ldots, F_{21}^{(p,p)}, F_{22}^{(1,1)}, F_{22}^{(1,2)}, \ldots, F_{22}^{(p,p)}, \ddot{\Lambda}_1^{(1)}, \ldots, \ddot{\Lambda}_1^{(j)}]^T$ and $\hat{\ddot{\theta}}_C = [\hat{F}_1^{(1)}, \ldots, \hat{F}_1^{(p)}, \hat{F}_{11}^{(1,1)}, \hat{F}_{11}^{(1,2)}, \ldots, \hat{F}_{11}^{(p,p)}, \hat{F}_{12}^{(1,1)}, \hat{F}_{12}^{(1,2)}, \ldots, \hat{F}_{12}^{(p,p)}, \hat{F}_{21}^{(1,1)}, \hat{F}_{21}^{(1,2)}, \ldots, \hat{F}_{21}^{(p,p)}, \hat{F}_{22}^{(1,1)}, \hat{F}_{22}^{(1,2)}, \ldots, \hat{F}_{22}^{(p,p)}, \hat{\ddot{\Lambda}}_1^{(1)}, \ldots, \hat{\ddot{\Lambda}}_1^{(j)}]^T$. Using results from Cheng, Fine and Kosorok (2007) it can be shown that

$$\sqrt{n}[\hat{\ddot{\theta}}_C - \ddot{\theta}_C] \rightsquigarrow Z_{\ddot{\theta}_C},$$

where $Z_{\ddot{\theta}_C} = [Z_{F_{11}}, \ldots, Z_{F_{1p}}, Z_{F_{1111}}, Z_{F_{1112}}, \ldots, Z_{F_{11pp}}, Z_{F_{1211}}, Z_{F_{1212}}, \ldots, Z_{F_{12pp}}, Z_{F_{2111}}, Z_{F_{2112}}, \ldots, Z_{F_{21pp}}, Z_{F_{2211}}, Z_{F_{2212}}, \ldots, Z_{F_{22pp}}, Z_{\Lambda_1}, \ldots, Z_{\Lambda_p}]^T$ is a $2p + 3p^2$ dimensional mean 0 Gaussian process. The proofs for $\widehat{\ddot{CN}}$ and $\widehat{\ddot{CM}}$ follow in a similar manner to $\widehat{CN}$ and $\widehat{CM}$.

1. $\widehat{\ddot{CN}}$: The relevant mapping is for $\widehat{\ddot{CN}}$

$$\phi_{\ddot{CN}jj'}(\ddot{\theta}_C) = F_{11}^{(j,j')} - F_1^{(j)} F_1^{(j')}$$

   Using this mapping the desired results can be shown using similar methods to $\widehat{CN}$ in the proof of Theorem 1 above.

2. $\widehat{\ddot{CM}}$: The relevant mapping for $\widehat{\ddot{CM}}$ is

$$\begin{aligned}
&\phi_{\ddot{CM}jj'}(\ddot{\theta}_C)(t) = \\
&\ddot{\Lambda}_1^{(j)}(t)\ddot{\Lambda}_1^{(j')}(t)\ddot{S}^{(j,j')}(t,t) + \int_0^t \ddot{\Lambda}_1^{(j)}(s)\ddot{\Lambda}_1^{(j')}(s)F_{22}^{(j,j')}(ds) \\
&+ \int_0^t \int_0^t \{1-\ddot{\Lambda}_1^{(j)}(s_1)\}\{1-\ddot{\Lambda}_1^{(j')}(s_2)\}F_{11}^{(j,j')}(ds_1, ds_2) \\
&+ \int_0^t \{1-\ddot{\Lambda}_1^{(j)}(s)\}\{-\ddot{\Lambda}_1^{(j')}(t)\}\{F_1^{(j)}(ds) - F_{11}^{(j,j')}(ds,t) - F_{12}^{(j,j')}(ds,t)\} \\
&+ \int_0^t \{1-\ddot{\Lambda}_1^{(j')}(s)\}\{-\ddot{\Lambda}_1^{(j)}(t)\}\{F_1^{(j')}(ds) - F_{11}^{(j,j')}(t,ds) - F_{21}^{(j,j')}(t,ds)\} \\
&+ \int_0^t \int_{s_1}^t \{1-\ddot{\Lambda}_1^{(j)}(s_1)\}\{-\ddot{\Lambda}_1^{(j')}(s_2)\}F_{12}^{(j,j')}(ds_1, ds_2) \\
&+ \int_0^t \int_{s_1}^t \{1-\ddot{\Lambda}_1^{(j')}(s_1)\}\{-\ddot{\Lambda}_1^{(j)}(s_2)\}F_{12}^{(j',j)}(ds_1, ds_2),
\end{aligned}$$

This mapping can be shown to be Hadamard differentiable in the direction of $Z_{\ddot{\theta}_C}$ through repeated application of Lemma 12.3 from Kosorok (2008) and Lemma A5 from Cheng, Fine and Kosorok (2007). The results then follow using similar methods to the proof for $\widehat{CM}$ in Theorem 1 above.

The results for $\widehat{\ddot{R}N}$ and $\widehat{\ddot{R}M}$ can be obtained by defining $\ddot{\theta}_R = [F_1^{(1)}, \ldots, F_1^{(p)}, \ddot{CN}^{(1,1)}, \ddot{CN}^{(1,2)}, \ldots, \ddot{CN}^{(p,p)}, \ddot{CM}^{(1,1)}, \ddot{CM}^{(1,2)}, \ldots, \ddot{CM}^{(p,p)}]^T$ and $\hat{\ddot{\theta}}_R = [\hat{F}_1^{(1)}, \ldots, \hat{F}_1^{(p)}, \widehat{\ddot{CN}}^{(1,1)}, \widehat{\ddot{CN}}^{(1,2)}, \ldots, \widehat{\ddot{CN}}^{(p,p)}, \widehat{\ddot{CM}}^{(1,1)}, \widehat{\ddot{CM}}^{(1,2)}, \ldots, \widehat{\ddot{CM}}^{(p,p)}]^T$. Using the results from above

$$\sqrt{n}[\hat{\ddot{\theta}}_R - \ddot{\theta}_R] \rightsquigarrow Z_{\ddot{\theta}_R},$$

where $Z_{\ddot{\theta}_R} = [Z_{F11}, \ldots, Z_{F1p}, Z_{\ddot{CN}11}, Z_{\ddot{CN}12}, \ldots, Z_{\ddot{CN}pp}, Z_{\ddot{CM}11}, Z_{\ddot{CM}12}, \ldots, Z_{\ddot{CM}pp}]^T$ is a $p + 2p^2$ dimensional mean 0 Gaussian process. The results for $\widehat{\ddot{R}N}$ and and $\widehat{\ddot{R}M}$ can be shown using similar methods to $\widehat{RN}$ and $\widehat{RM}$ above using the two mappings below.

1. $\widehat{\ddot{R}N}$: The relevant mapping is

$$\phi_{\ddot{RN}jj'}(\ddot{\theta}_R) = \frac{\ddot{CN}^{(j,j')}}{\sqrt{F_1^{(j)}(1 - F_1^{(j)})}\sqrt{F_1^{(j')}(1 - F_1^{(j')})}}.$$

2. $\widehat{\ddot{R}M}$: The relevant mapping is

$$\phi_{\ddot{RM}jj'}(\ddot{\theta}_R) = \frac{\ddot{CM}^{(j,j')}}{\sqrt{F_1^{(j)}}\sqrt{F_1^{(j')}}}.$$

Proof of Theorem 3. The consistency of $\hat{V}$ can be established through a straightforward application of the Davis-Kahan Theorem, and the consistency of $\hat{\Xi}$ can be established through Weyl's inequality. From here asymptotic normality can be established using steps similar to Theorem 13.5.1 in Anderson (2003), which derives the asymptotic variances of the eigenvectors and eigenvalues for the sample covariance matrix when data have a multivariate normal distribution and the sample covariance matrix has wishart distribution. However the steps can be straightforwardly extended to any asymptotically normal positive definite estimate of the covariance or correlation matrix as shown below.

Consider the transformation $Q = V^T \widehat{\Sigma} V$. Then by the Delta method $\sqrt{n}(Q - \Xi) \rightarrow_d N(0, J_V \Psi_\Sigma J_V^T)$ where $J_V = V^T \otimes V^T$. It can be shown that $\widehat{\Sigma}$ and Q have

the same singular values so Q can be represented as

$$Q = G\widehat{\Xi}G^T, \tag{A.2}$$

for orthogonal G. G can be uniquely defined with the constraint $g_{ii} \geq 0$. Note that $\widehat{\Xi} = \hat{V}^T\widehat{\Sigma}\hat{V}$, which together with the fact that $V^T\hat{V}$ is orthogonal implies that every column of G is equal to $\pm$ the corresponding column of $V^T\hat{V}$. Because $\hat{V} \to V$ the constraint that $g_{ii} \geq 0$ implies that $G \to I$ where I the identity matrix. Define $U = \sqrt{n}(Q - \Xi)$, $D = \sqrt{n}(\hat{\Xi} - \Xi)$ and $W = \sqrt{n}(Y - I)$. When we combine Equation (A.2) with $GG^T = I$, and the conditions $g_{ii} > 0$ and $\hat{\xi}_1, > \ldots, > \hat{\xi}_p$, we get a set of one to one functions from Q to G and $\hat{\Xi}$ except on a set of measure zero which are continuously differentiable and have well defined inverses in the neighborhood of $\hat{\Xi} = \Xi$ and $G = I$. Therefore by the fact that U is asymptotically normal with mean zero, from the delta method W and D are also asymptotically normal with mean zero. Further because the column of $\hat{V}$ is equal to $\pm$ the corresponding column of VG and $G \to I$, $\sqrt{n}(\hat{V} - V)$ has the same limiting distribution as $\sqrt{n}(VG - V)$ which is asymptotically normal with mean zero by the delta method.

In order to show that the limiting variances are functions of Ψ_Σ, V, and Λ, using similar algebra from Theorem 13.5.1 in (Anderson (2003)) we get the following equalities

$$U = W\Lambda + D + \Lambda W^T + o_p(1) \tag{A.3}$$

$$0 = W + W^T + o_p(1) \tag{A.4}$$

By combining results from Equations (A.3) and (A.4) and ignoring the $o_p(1)$ terms we obtain

$$w_{ii} = 0 \tag{A.5}$$

$$d_{ii} = u_{ii} \tag{A.6}$$

$$w_{ij} = \frac{u_{ij}}{\lambda_j - \lambda_i} \text{ for } i \neq j \tag{A.7}$$

The limiting distribution for W and D can be solved for using the limiting distribution of U. The limiting distribution of W can be used to find the limiting distribution of $\sqrt{n}(\hat{V} - V)$.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd Edition. Wiley.

Bitouzé, D., Laurent, B. and Massart, P. (1999). A Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan–Meier estimator. *Annales de l'Institut Henri Poincare (B) Probability and Statistics* **35**, 735–763.

Cheng, Y., Fine, J. P. and Kosorok, M. R. (2007). Nonparametric association analysis of bivariate competing-risks data. *Journal of the American Statistical Association* **102**, 1407–1415.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *The Annals of Statistics* **16**, 1475–1489.

Fine, J. P., Jiang, H. and Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907–919.

Gill, R. D., van der Laan, M. J. and Wellner, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. *Annales de l'IHP Probabilités et Statistiques* **31**, 545–597.

Gillespie, B., Gillespie, J. and Iglewicz, B. (1992). A comparison of the bias in four versions of the product-limit estimator. *Biometrika* **79**, 149–155.

Jazić, I., Schrag, D., Sargent, D. J. and Haneuse, S. (2016). Beyond composite endpoints analysis: Semicompeting risks as an underutilized framework for cancer research. *JNCI: Journal of the National Cancer Institute* **108**. Web: `https://doi.org/10.1093/jnci/djw257`.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Hoboken.

Karnofsky, D. A., Abelmann, W. H., Craver, L. F. and Burchenal, J. H. (1948). The use of the nitrogen mustards in the palliative treatment of carcinoma. with particular reference to bronchogenic carcinoma. *Cancer* **1**, 634–656.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

Lin, D. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, 573–581.

Lin, D. Y. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine* **16**, 901–910.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 559–572.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, 495–512.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics–Theory and Methods* **22**, 965–984.

van der Laan, M. J. (1993). *Modified EM-estimator of the Bivariate Survival Function*. Rijksuniversiteit Utrecht. Mathematisch Instituut.

Von Hoff, D. D., Ervin, T., Arena, F. P., Chiorean, E. G., Infante, J., Moore, M. et al. (2013). Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *New England Journal of Medicine* **369**, 1691–1703.

Benjamin W. Langworthy

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA.

E-mail: langworthy.ben@gmail.com

Jianwen Cai

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA.

E-mail: cai@bios.unc.edu

Robert W. Corty

School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA.

E-mail: rcorty@gmail.com

Michael R. Kosorok

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

E-mail: kosorok@bios.unc.edu

Jason P. Fine

School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

E-mail: jpfine1606@yahoo.com

(Received March 2021; accepted November 2021)